

## Relationship between Amino Acid Sequence and Secondary Structures of Proteins in Plants and Cereals

Shela GORINSTEIN

Department of Pharmaceutical Chemistry, School of Pharmacy, The Hebrew University of Jerusalem, P.O.B. 12065, Jerusalem, Israel

Received September 7, 1992

Alcohol-soluble proteins from *Amaranthus* seeds differ from prolamins fractions of cereals and plants. Albumins and globulins are the major fractions of *Amaranthus*. Peptide sequences of storage and membrane proteins from different cereals and plants were compared. Secondary structure of proteins was computerized by dot matrix and hydropathy analyses.

The major storage proteins in most cereals (wheat, barley, rye, maize, and sorghum) are alcohol-soluble prolamins, but such prolamins are not characteristic of all plants.<sup>1-3</sup> Oats and rice accumulate globulins and glutelins respectively as their major storage proteins. Oat globulin and rice glutelin are highly homologous and are related to the legume 11S storage proteins.<sup>4-7</sup> Seeds of grain amaranth contain a large amount of albumin and globulin and traces of prolamins.<sup>3,7,8</sup> The amino acid composition of amaranth globulin was almost the same as those of soybean and oat globulins,<sup>8-10</sup> but relative amounts of proteins distributed in different protein fractions were very close to buckwheat<sup>8</sup> as well as to triticale (germ) and corn opaque-2.<sup>3</sup> The only data available in the literature about amino acid sequences in amaranth plants relate to membrane and herbicide binding proteins of the chloroplast.<sup>11,12</sup> Albumins and globulins according to their amino acid composition are close to soybean proteins.<sup>8,9</sup> The comparison of sequences of amino acids may be done assuming that alcohol-soluble proteins in amaranth are not the main protein fractions.<sup>3,8,9</sup>

This paper deals with the extraction and separation of alcohol-soluble proteins in amaranth seeds. Based on the sequence of storage and membrane proteins in soybean<sup>13</sup> and amaranth<sup>11</sup> we searched for similarity in other plants, and for their secondary structure.

### Materials and Methods

**Sample preparation.** Whole mature seeds of amaranth: *Amaranthus* (*A.*) *cruentus* (purple); *A. flavus*; *A. caudatus*; *A. hypochondriacus*; and *A. cruentus* (yellow) were investigated. These plants were donated by Dr. Alirindo Moreira Sales, Instituto De Tecnologia De Alimentos, Campinas, Brazil. Cereals such as high tannin sorghum, normal maize, and oats, which were used for comparison of prolamins were taken from the Plant Breeding Laboratory, Sementes, Agronegocios, Brazil.

Amaranth and other seeds were ground in a mill with a 60-mesh screen and defatted in a Soxhlet extractor with *n*-hexane for 10 h. The meal was stored at 5°C after removal of hexane.

**Protein extraction.** Proteins were extracted stepwise following the method of Landry and Moureaux.<sup>14</sup> Extractibility of alcohol-soluble proteins from amaranth seeds was studied at 20°C using isopropanol (2-ProOH) mixtures with 2-mercaptoethanol (2-ME) varying from 0 to 5% with changes in extraction time, concentration of reducing agent, and proportion of solvent to solid.<sup>9,10</sup> Prolamins of all plants were extracted

with 55% 2-ProOH and 5% 2-ME (optimum extraction conditions, which were found in previous studies)<sup>9,10</sup>

**Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).** SDS-PAGE was done by the procedure of Laemmli.<sup>15</sup> Extracts were combined, lyophilized, and dissolved in sample buffer which contained 10% glycerol, 5% 2-ME, and 2% sodium dodecyl sulfate (SDS) in 0.125 M tris(hydroxymethyl)aminomethane (Tris-HCl), pH 6.8. Then the extracts were boiled for 5 min before being put on the gel. Proteins were then precipitated with acetone (1:2 volumes) at -20°C overnight, and the precipitate was dissolved in the same sample buffer. The gels were 1.5 mm thick and consisted of a 2 cm stacking gel and a 10 cm running gel. The (5-20%) polyacrylamide gel (PAAG) gradients were made from stock solutions of 0% and 36% acrylamide in 0.8% *N,N'*-methylene-bis-AA (BIS) and 0.1% SDS in 0.375 M Tris-HCl, pH 8.8. Fifty µg of protein was put in the sample slots. Electrophoresis was done at 100 V for 4 h. Gels were stained for 2 h with 0.25% Coomassie Brilliant Blue R in methanol-water-acetic acid (5:5:1, v/v) and destained in the same solvent. Molecular weight (MW) standards were used to estimate protein subunit molecular weights.<sup>16</sup>

**Computer analysis of amino acid sequences of proteins from plants and cereals.** The amino acid sequences of storage proteins that were stored in a data bank were compared with those of other proteins by computer analysis using the FASTA and TFASTA programs, based on the parameters and algorithm of known sequences,<sup>17,18</sup> dot matrix analyses,<sup>19</sup> and Basic Local Alignment Search Tool (BLAST).<sup>20</sup> These programs were installed in the VAX/VMS operating system, and in an IBM PC microcomputer and available at the Computer Center at the Hebrew University.<sup>19</sup>

### Results and Discussion

#### Results of electrophoretic separation

Based on the solvent systems used for extraction of alcohol-soluble proteins from amaranth and other seeds,<sup>9</sup> their separation gave the following results (Fig. 1). Position A of Fig. 1 shows alcohol-soluble proteins of amaranth seed extracted with 55% 2-ProOH-5% 2-ME (6:1, v/w) and position B of Fig. 1 presents the same five types of amaranth compared to oats, maize, and sorghum, extracted with 55% 2-ProOH-5% 2-ME (10:1, v/w). These proteins contain 80-85% of polypeptides of 10-14 kDa and 7% of 20-kDa polypeptides, the rest being minor fractions. Only slight differences were observed in subunits of five amaranth species. As can be seen from Fig. 1, alcohol-soluble proteins of amaranth extracted with different solvents (which are

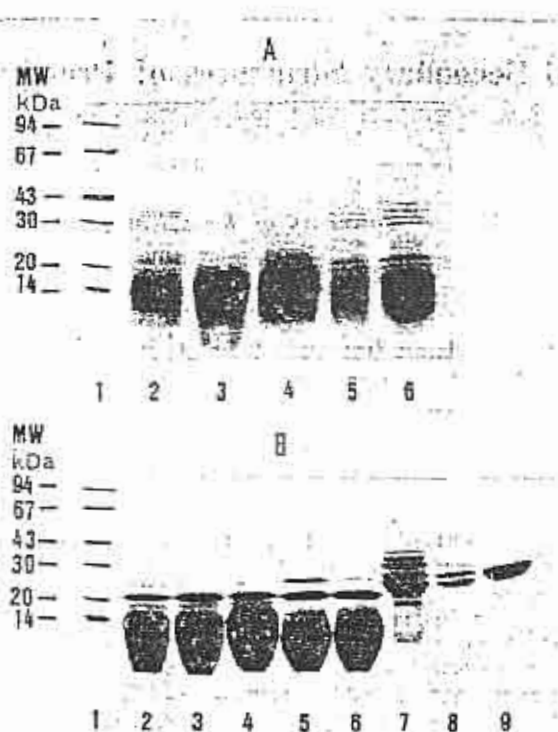


Fig. 1. SDS-PAGE of Alcohol-soluble Proteins from Different Species of *Amaranthus*.

Seeds and cereals in gradient (5–20%) PAAG. Proteins were extracted from the meal with:

A. 55% 2-ProOH–5% 2-ME (6:1, v/v). 1—marker; 2, 3, 4, 5, 6—*A. cruentus* (purple); *A. flavus*; *A. caudatus*; *A. hypochondriacus*; *A. cruentus* (yellow), respectively.

B. 55% 2-ProOH–5% 2-ME (10:1, v/v). 1—marker; 2, 3, 4, 5, 6, 7, 8, 9—*A. cruentus* (purple); *A. flavus*; *A. caudatus*; *A. hypochondriacus*; *A. cruentus* (yellow); oats, normal maize, and high-tannin sorghum, respectively. Proteins were stained with Coomassie Brilliant Blue. Molecular mass markers were phosphorylase b (94 kDa), hemoglobin canine (67 kDa), ovalbumin (43 kDa), carboanhydrase (30 kDa), trypsin inhibitor, soybean (20 kDa), and lactalbumin from bovine milk (14 kDa).

not shown in our previous work<sup>9)</sup> have nonseparated subunits in the region of 8–14 kDa. These protein subunits differ completely from those of other cereals such as oats, maize, and high-tannin sorghum. Prolamins extracted from oats, maize, and sorghum did not show any electrophoretic relationship with amaranth alcohol-soluble fractions. The data proved that alcohol-soluble proteins can't be the main protein fraction in cereal-like plants like amaranth.

*Graphic matrix analysis of protein sequences. Rapid similarity searches of nucleic acids and protein data bank. Homology of amino acid sequences*

Using the FASTA and TFASTA computer programs, homology was detected between a short region of 29 residues of legumes and other plants. The search for parts of the polypeptide chains which form the structural regions was done in two ways: comparison of the sequences between various plants to detect identity, and comparison of the identical sequence found with one of them to detect repeats.

*Glycine max* (soybean) sequences of 12S globulin subunit, which were the basis of our search, showed homology (Table I) of 100.0%, 93.1%, and 89.7% in chains of G<sub>2</sub> (2) and A2B1a; G<sub>1</sub> (1) and Ala; and G<sub>3</sub> (3).<sup>4,13,21–25)</sup> Similar homology between legumes and other plants has been noted previously by Borroto and Dure.<sup>21)</sup> In *Vicia faba* (fava bean) of garden and field beans the sequences of  $\alpha$ -chain were 72.4% and 71.4%. The sequences in the  $\beta$ -chain of legumin from *Pisum sativum* (garden pea) showed homology of 69.0%. It was also observed that globulin of *Arabidopsis thaliana* and *Brassica napus* (rape); storage protein of *Cucurbita maxima* (pumpkin) and soybean (A3–B4); globulins of *Helianthus annuus* (common sunflower) and fava bean  $\beta$ -chain had respectively the same homology of 46.4%; 55.2%, and 51.7% in the sequences of 29 overlaps.

Oryzerin and glutelin of *Oryza sativa* (rice); globulin of *Gossypium hirsutum* (upland cotton), and *Avena sativa* (oats) showed the following homology as 62.1%; 58.6%; 48.3%,

Table I. Amino Acid Homology of Globulin Proteins in 29 Amino Acid Overlaps

Plants	Sequence		Homology (%)
	10	20	
Soybean*	NGIDETICTMRLRQNIQGHSSPDIYNPQA		100.0
Soybean**	NGIDETICTMRLRHNIQTSSPDIYNPQA		93.1
Soybean G <sub>3</sub> (3)	NGIDETICTMRLRHNIQTSSPDIYNPQA		89.7
Fava bean ( $\alpha$ )	NGLEETVCTAKLRLNIGSSSPDIYNPQA		72.4
Field bean ( $\alpha$ )	GLEETVCTAKLRLNIGSSSPDIYNPQA		71.4*
Garden pea ( $\beta$ )	NGLEETVCTAKLRLNIGPSSSPDIYNPEA		69.0
Rice (oryzerin)	NGLDETFCTMRVVRQNIQDNPNRADTYNPPRA		62.1
Field bean ( $\beta$ )	GLEETICSLKIRENIAQPARADLYNPPRA		60.7*
Rice (glutelin)	NGLDETFCTLRVRQNIQDNPNRADTYNPPRA		58.6
Soybean (A3–B4)	NGVEENICTLKLHENIARPSRADFYNPQA		55.2
Cucurbit ( $\beta$ )	NGLEETICTLRLKQNIQGRSVRADVFNPPRA		55.2
Sunflower	NGVEETICSMKFKVNIQDNPSQADFYNPQA		51.7
Fava bean ( $\beta$ )	NGLEETICSLKIRENIAQPARADLYNPPRA		51.7
Cotton	NGLEETFCNRIKENLADPERADIFNPQA		48.3
Arabidopsis	NGLEETLCTNRCTENLQOPSDADVYKPSL		46.4*
Rape	NGLEETLCTMRCTENLDDPSSADVYKPSL		46.4
Oats	NGLEENFCSLEARQNIENPKRADTYNPPRA		44.8

\* A2B1a, gene encoding a glycinin A2B1a subunit precursor<sup>21)</sup>; G<sub>2</sub> (2), glycinin G<sub>2</sub> (2) gene<sup>24)</sup>.

\*\* G<sub>1</sub> (1), glycinin G<sub>1</sub> (1) gene<sup>22)</sup>; A-1a, glycinin A-1a subunit precursor.<sup>25)</sup>

\* 28 overlaps.

and 44.8% in the sequences in 29 overlaps in comparison with soybean globulin. Earlier work by others proved that the rice glutelin basic subunit showed homology to pea legumin and other 11S legumin-like storage proteins.<sup>1,4,5,26,27</sup> Conservation of peptide sequences among the globulin-type storage proteins from several legumes and cereals has been reported.<sup>28-30</sup> Despite their different solubility properties, rice glutelin shares many biochemical and cellular properties with 11S legume storage proteins such as pea legumin. Both types of proteins are synthesized as larger precursors on rough endoplasmic reticulum membranes, transported and packaged into protein bodies via the Golgi complex, and after translation are proteolyzed,

forming acidic and basic subunits.<sup>3,28</sup> Within the results of the search shown in Table I, the identity of globulins in the plant *Amaranthus* was not found. Based on sequencing of the 32k thylakoid membrane protein precursor from chloroplasts of *Amaranthus hybridus*.<sup>22</sup> The search by FASTA and TFASTA showed 93.5% homology in 475 overlaps of *Pisum sativum*, *Gossypium hirsutum*, and *Oryza sativa*. *A. hybridus*<sup>31,32</sup> showed also around 99.7% homology in 353 overlaps with cotton and soybean (Table II).

As can be seen from the data of Table II, even the membrane protein of amaranth showed near homology to the same plants as in storage globulin proteins (Table I).

The BLAST<sup>20,33</sup> search based on the 32k thylakoid membrane protein from *Amaranthus hybridus* (99 and 119 amino acids) showed some matching only with 11S globulin

Table II. Identity of Membrane Proteins in 353 Amino Acid Overlaps

Plants	Homology (%)
<i>Amaranthus hybridus</i>	100.0
Cotton	99.7
<i>Glycine max</i> (soybean)	99.2
<i>Brassica napus</i> (rape)	99.8
<i>Pisum sativum</i> (garden pea)	98.6
<i>Secale cereale</i> (rye)	98.3
<i>Oryza sativa</i> (rice)	95.6
<i>Hordeum vulgare</i> (barley)	98.0
<i>Vicia faba</i> (broad bean)	96.0

Table III. Data Found by Blast When Searching Various Proteins in the Swissprot Database

Plant	Sequence	Parameters
<i>A. Hybridus</i> 99	AASVDEWLYNGGPYELIVLHF 119	Score = 52; Expect = 16
Pumpkin	161 PAGVSHWMYNRGQSDLVLIVF 181	Poisson P = 1.0; Length = 21

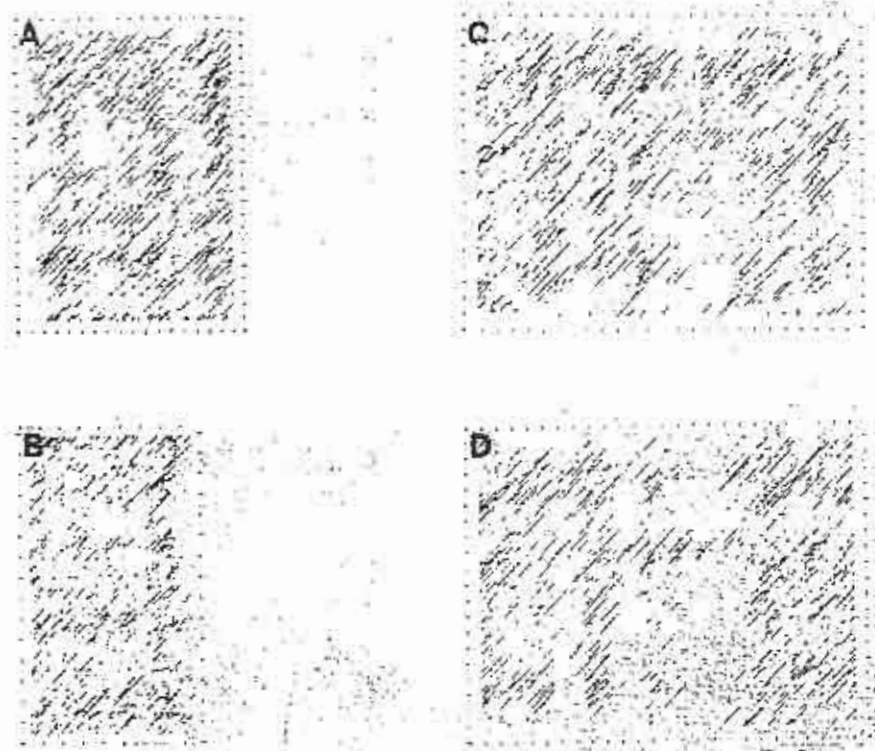


Fig. 2. Dot Matrix Analyses of Proteins from *Amaranthus* and Other Plants.

X axis: A, B—32k thylakoid membrane protein precursor of chloroplast from *A. hybridus* [1-353 amino acid (aa) overlaps]; C, D—storage proteins of oats and soybean; Y axis: A, B—11S globulin  $\beta$  subunit precursor of *Cucurbita maxima* (pumpkin), cotton chloroplast; C, D—membrane protein precursor of chloroplast *A. hybridus* (1-395 aa overlaps).

from *Cucurbita maxima* (pumpkin) which was done between the sequence of 161 and 181 amino acids.<sup>29</sup> This correspondence is shown in Table III.

Comparison of these two protein sequences created a file of the points of similarity between them in vertical (pumpkin) and horizontal (*A. hybridus*) dimensions. These comparisons were plotted with a DotPlot (Fig. 2, position A). DotPlot is the second part of a two-part set of programmes that generate dot-plots of the points of similarity between two sequences.<sup>18,19</sup> Nucleotide sequences were translated into peptide sequences and shown in Fig. 2, position B. Positions C and D of Fig. 2 show the comparison of *A. hybridus* translated into a peptide sequence with storage proteins of oats and soybean *G<sub>1</sub>(3)* from Table I.

Comparisons between the following pairs were done on Fig. 3.

Position (A): chloroplasts of *A. hybridus* and of *Oryza*

*sativa* with 98.6% of homology in 353 overlaps.

Position (B): chloroplasts of *A. hybridus* and of *Pisum sativum* with 98.6% of homology.

Position (C): chloroplasts of *A. hybridus* and *Glycine max* with 99.2% of homology.

Position (D): chloroplasts of *Amaranthus hybridus* and of *Gossypium hirsutum* which showed 93.1% homology in 475 overlaps.

Position (E): chloroplasts of *Amaranthus hybridus* and of *Oryza sativa* which showed 92.2% in 475 overlaps.

Position (F): chloroplasts of *Amaranthus hybridus* and of *Pisum sativum* which showed 93.5% homology in 475 overlaps. Since the diagonal line of homology is much more evident (Fig. 3), in this comparison, it is used throughout to compare the peptide and nucleotide sequences of the identical regions for investigated storage and membrane proteins. All comparisons shown in Fig. 3 visualize a diagonal of the surface. Considerable background is

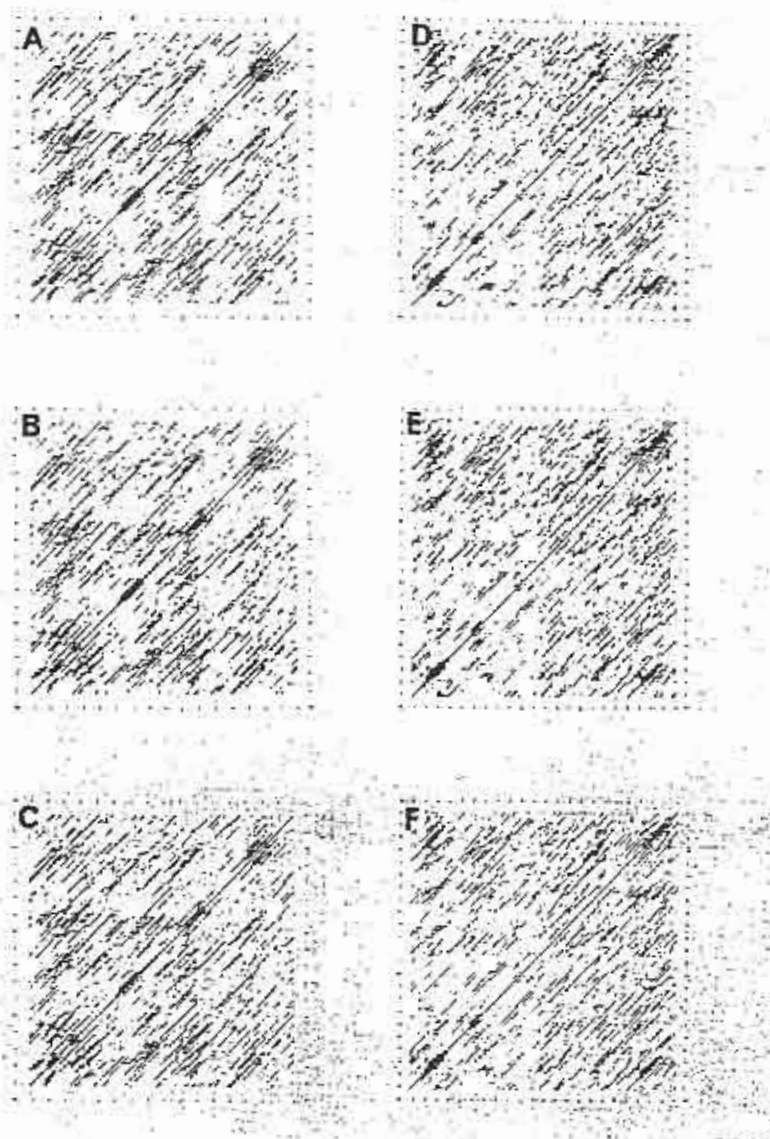


Fig. 3. Comparison of 32k Thylakoid Membrane Protein Precursor of Chloroplast from *Amaranthus hybridus* with Other Plants and Cereals.

X axis: A, B, C—membrane protein precursor of chloroplast from *A. hybridus* (1-353 aa overlaps); D, E, F—membrane protein precursor of chloroplast from *A. hybridus* (1-475 aa overlaps);

Y axis: A, B, C—storage proteins from (1-353 aa overlaps) rice, pea, soybean; D, E, F—storage proteins from (1-475 aa overlaps) cotton, rice, pea.

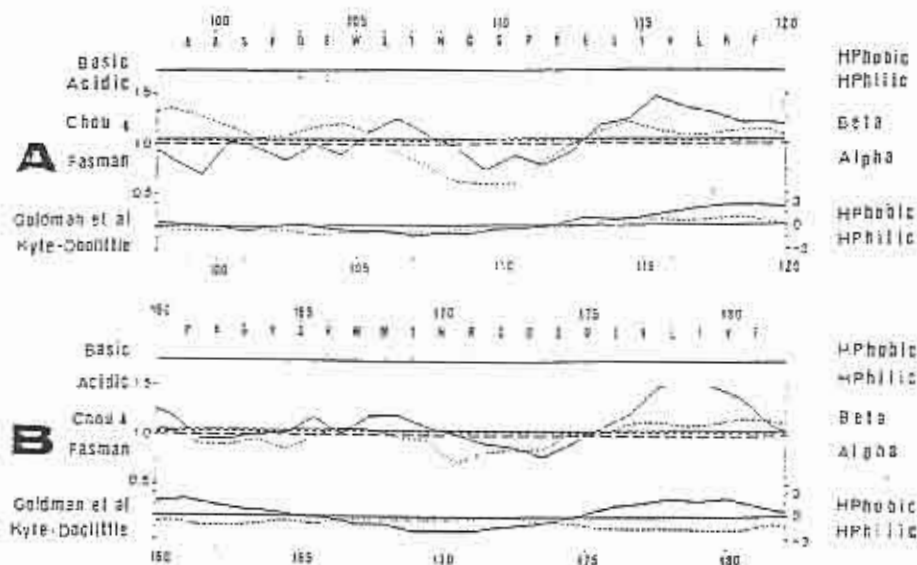


Fig. 4. Comparison of the Hydrophobicity Profiles and Secondary Structures Predicted from the Amino Acid Sequences of the 32 k Thylakoid Membrane Protein Precursor of Chloroplast (99 to 119aa overlaps) from *Amaranthus hybridus* and 11S Globulin  $\beta$  Subunit Precursor (161–181aa overlaps) of *Cucurbita Maxima* (B).

(---),  $\alpha$ -helices; (—),  $\beta$ -sheets.

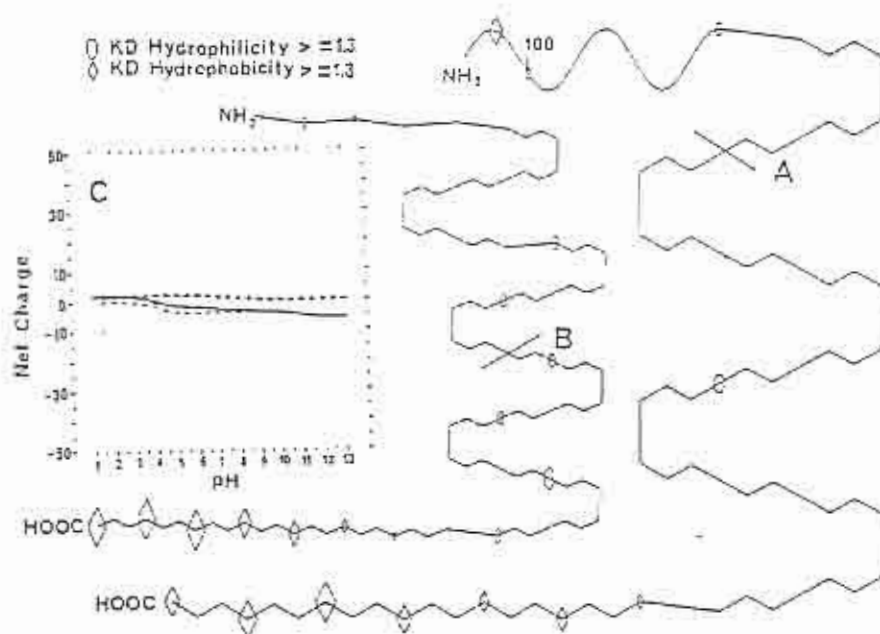


Fig. 5. Plot Structure of (A) 32 k Thylakoid Membrane Protein Precursor of Chloroplast from *Amaranthus hybridus*; (B) 11S Globulin  $\beta$  Subunit from *Cucurbita maxima*; (C) Isoelectric Plot of Peptide Sequence, Where (+) Positively Charged Residues and (–) Negatively Charged Residues, of 32 k Thylakoid Membrane Protein Precursor of Chloroplast from *Amaranthus hybridus*.

generated, although a diagonal line of homology throughout is discernible. Figure 2 shows a completely different picture. Some matching can be seen in the position of Fig. 2, and tandem repeated sequences occurring in both proteins appear as blocks of dots. In spite of the lack of homology between the membrane and storage proteins there are indications of some matching.

*Protein analysis. Secondary structure, hydrophobicity, isoelectric plots, peptide plots, and moment*

A peptide plot (Fig. 4A) measured the protein secondary structure and hydrophobicity in parallel panels of the same plot in the 32 k thylakoid membrane protein precursor of

the chloroplast [99 to 119 amino acid (aa) overlaps] from *A. hybridus*. In the same length of 21 amino acids a peptide plot was obtained for 11S globulin  $\beta$  subunit precursor (161–181 amino acid overlaps) from *Cucurbita maxima*, Fig. 4B. The first line in Fig. 4 (Positions A and B) shows the sequence itself for 21 residues of each protein. The second line represents the position in the sequence of basic to hydrophobic (top), and acidic to hydrophilic amino acids. The third line measures  $\alpha$ -helix and  $\beta$ -sheet.<sup>34</sup> A Kyte-Doolittle curve<sup>35</sup> shows (Fig. 4) the average of a residue-specific hydrophobicity index over a window of nine residues with the hydrophobic region in the upper half and hydrophilic in the lower half of the axis.<sup>35</sup> Goldman,

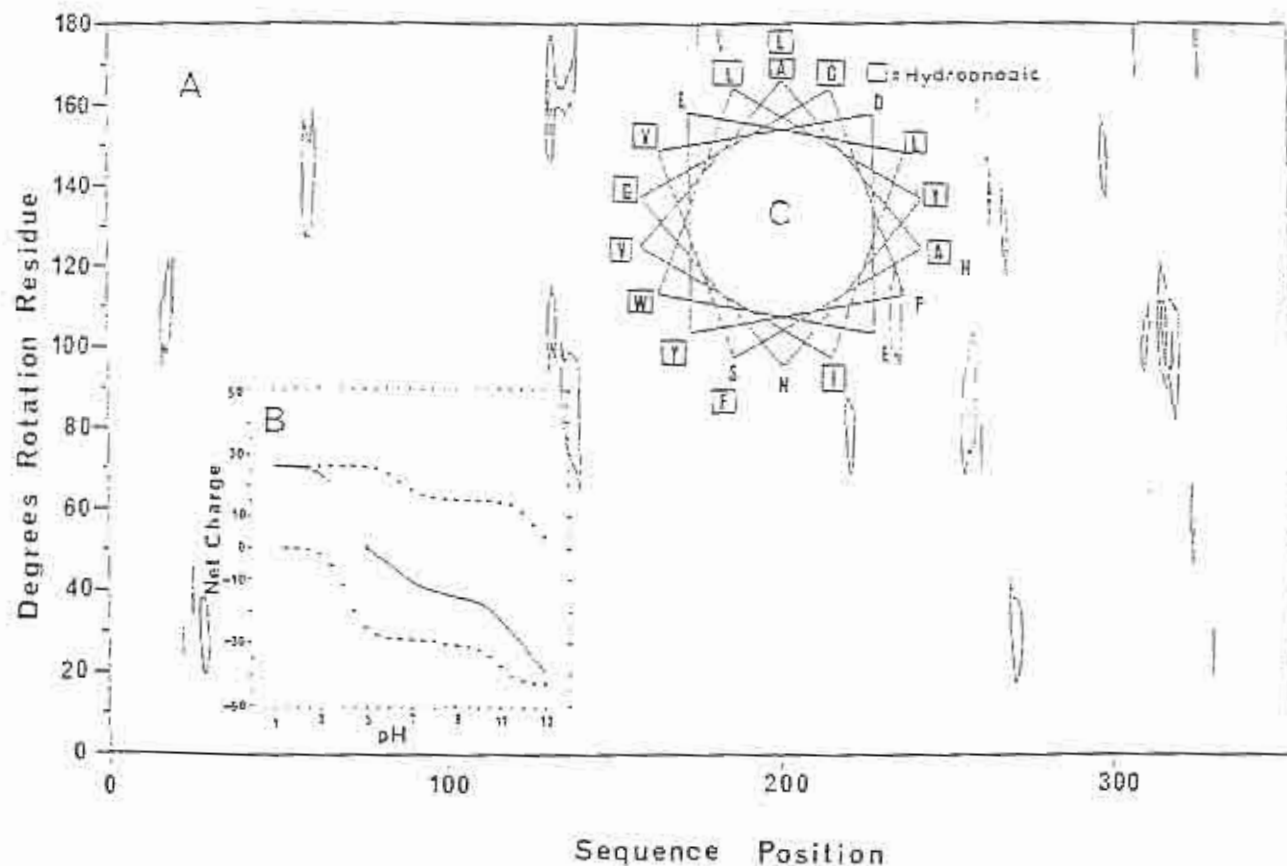


Fig. 6. Contour Plot of the Helical Hydrophobic Moment of a Peptide Sequence (A), Isoelectric Plot of the Total Positive and Negative Charges and the Net Charge as a Function of pH (B) (1 to 353 aa overlaps). Hydrophobicity of a Peptide-helical Wheel (C) of 32k Thylakoid Membrane Protein Precursor of Chloroplast (1 to 119 aa overlaps) from *Amaranthus hybridus*.

Engelman, and Steiz curve identifies nonpolar transbilayer helices with hydrophobic regions in the upper half and hydrophilic ones in the lower one.<sup>36)</sup> As may be seen from Fig. 4 the two proteins in their secondary structure have little similarity in these specific 21 residues.

Peptide structures were calculated for the same two proteins and presented as plot structures. A plot was done according to the Chou-Fasman<sup>34)</sup> prediction system. Secondary structure predictions for 32k thylakoid membrane protein precursor of chloroplast from amaranth (Fig. 5A) differed from the 11S globulin beta subunit from pumpkin (Fig. 5B). It might be expected that amaranth and pumpkin, which have close identity in certain regions, show also the greatest similarities in the pattern of predicted Chou-Fasman structure. The secondary structure prediction corresponds with the regions of high identity within the chains of amino acid residues.

An isoelectric plot of the total positive and negative charges and the net charge as a function of pH for the 32k peptide sequence of thylakoid membrane protein precursor is shown in Fig. 5C. The net change of a protein is calculated as the sum of the number of positively charged residues (protonated lysine, arginine, histidine), minus the number of negatively charged residues (deprotonated tyrosine, cysteine, glutamate, aspartate), plus the number of protonated amino termini, minus the number of deprotonated carboxyl termini. This plot was compared with 11S globulin  $\beta$  subunit precursor from *Cucurbita maxima* (Fig. 5D). The difference was in charged groups as well

as in isoelectric points ( $pI$ ). The  $pI$  for all sequences was 5.05 (Fig. 6B). Protein of *Amaranthus* in 21 residues has a  $pI$  of 4.00 with the following charged groups: His-1; Tyr-2; Glu-2 and Asp-1. *Cucurbita maxima* protein has a  $pI$  of 7.53 with Arg-1; His-1; Tyr-1 and Asp-1.

The hydrophobic moment is the hydrophobicity of a peptide measured for different angles of rotation per residue. The hydrophobic moment is calculated for all angles of rotation from 0 to 180 degrees. Moment helps to recognize "amphiphilic" structures by identifying when the residues on one side of the structure are more hydrophobic than on the other. It is a measure of the probability that the peptide at any particular position is located at the interface between the interior of the protein and the surface or more exactly, that the peptide separates hydrophobic and hydrophilic regions. Figure 6A shows a contour plot of the helical hydrophobic moment of 32k the thylakoid membrane protein precursor of chloroplast from amaranth. A helical wheel (Fig. 6C) of the 32k protein showed 14 hydrophobic regions in comparison with the helical wheel of *Cucurbita maxima* of 13.

It was found through the sequence analyses presented here that soybean legumin (100% homology), the sequence of which is known, has vestigial sequence homology with other legume storage proteins and globulins of cereals. Two sequences from amaranth and pumpkin in their membrane and globulin proteins were also compared. Secondary structures were predicted for these proteins.

*Acknowledgment.* The author is grateful to Miss I. Perlov for her valuable technical assistance.

## References

- 1) T. W. Okita, H. B. Krishnan, and W. T. Kim, *Plant Sci.*, **57**, 103 (1988).
- 2) K. Berroto and L. Duré, III, *Plant Molec. Biol.*, **8**, 113-131 (1987).
- 3) R. Bressani and L. A. Garcia-Vela, *J. Agric. Food Chem.*, **38**, 1205-1209 (1990).
- 4) W. Higuchi and C. Fukazawa, *Gene*, **55**, 245-253 (1987).
- 5) K. Abe, Y. Emori, H. Kawasaki, H. Kondo, K. Suzuki, and S. Arai, *Agric. Biol. Chem.*, **53**, 2069-2073 (1989).
- 6) T. W. Okita, Y. S. Hwang, J. Hnijo, W. T. Kim, A. P. Aryan, R. Larson, and H. B. Krishnan, *J. Biol. Chem.*, **264**, 12573-12581 (1989).
- 7) M. A. Shotwell, S. K. Boger, R. S. Chesnut, and B. A. Larkins, *J. Biol. Chem.*, **265**, 9652 (1990).
- 8) Y. Konishi, Y. Fumita, I. Ikeda, K. Okino, and H. Fuwa, *Agric. Biol. Chem.*, **49**, 1453-1459 (1985).
- 9) S. Gorinstein, I. Arnao de Nue, and P. Arruda, *J. Agric. Food Chem.*, **39**, 848 (1991).
- 10) S. Gorinstein, R. Moshe, L. J. Green, and P. Arruda, *J. Agric. Food Chem.*, **39**, 851 (1991).
- 11) J. Hirschberg and L. McIntosh, *Science*, **222**, 1346-1349 (1983).
- 12) C. B. Michalowski, H. J. Bohnert, D. F. Klessig, and J. O. Berry, *Nucl. Acids Res.*, **18**, 2187 (1990).
- 13) T. J. Cho and N. C. Nielsen, *Nucl. Acids Res.*, **17**, 4388 (1989).
- 14) J. Landry and T. Moureaux, *J. Agric. Food Chem.*, **28**, 1186-1191 (1980).
- 15) U. K. Laemmli, *Nature*, **227**, 680-685 (1970).
- 16) B. D. Plikayis, G. M. Carione, P. Edwards, W. Mayer, and W. Robust, *Anal. Biochem.*, **152**, 346-364 (1986).
- 17) D. J. Lipman and W. R. Peakson, *Science*, **227** (27), 1433-1441 (1985).
- 18) J. R. Devereux, Program Manual "Sequence Analysis Software Package GCG" of the Genetic Computer Group, Version 6.2, University Research Park, 575 Science Drive, Suite B, Madison, WI 53711. E-mail: HPLP@GCG. Com., modules 4-7, 10 (1989).
- 19) J. Maizel and R. Lenk, *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 7665-7669 (1981).
- 20) S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *J. Mol. Biol.*, **215**, 403 (1990).
- 21) C. Fukazawa, T. Momma, W. Higuchi, and K. Udaka, *Nucl. Acids Res.*, **15**, 3117 (1987).
- 22) N. C. Nielsen, C. D. Dickinson, T. J. Cho, V. H. Thanh, B. J. Scallion, R. L. Fischer, T. L. Sims, G. N. Drawns, and R. L. Goldberg, *Plant Cell*, **1**, 313-328 (1989).
- 23) T. L. Sims and R. B. Goldberg, *Nucl. Acids Res.*, **17**, 4386 (1989).
- 24) V. H. Thanh, N. E. Turner, and N. C. Nielsen, *Nucl. Acids Res.*, **17**, 4387 (1989).
- 25) T. Negoro, T. Momma, and C. Fukazawa, *Nucl. Acids Res.*, **13**, 6713-6731 (1985).
- 26) F. Takaiwa, S. Kikuchi, and K. Oono, *Nucl. Acids Res.*, **17**, 3289 (1989).
- 27) L. Wen, J. K. Huang, B. H. Johnson, and G. R. Reeck, *Nucl. Acids Res.*, **17**, 9470 (1989).
- 28) P. Plietz, B. Drescher, and G. Damaschun, *Inst. J. Biol. Macromol.*, **9**, 161-165 (1987).
- 29) M. Hayashi, H. Mori, M. Nishimuka, T. Akazawa, and I. Hara-Nishimuka, *Eur. J. Biochem.*, **172**, 627-632 (1988).
- 30) S. Sjoedahl, J. Roedin, and L. Rask, *Eur. J. Biochem.*, **196**, 617-621 (1991).
- 31) M. Keller and E. Stutz, *FEBS Lett.*, **175**, 173 (1984).
- 32) V. A. Dzelzkalns, G. C. Owens, and L. Bogorad, *Nucl. Acids Res.*, **12**, 8917 (1984).
- 33) S. F. Altschul, *J. Mol. Biol.*, **219**, 555 (1991).
- 34) P. G. Chou and G. D. Fasman, *Adv. in Enzymol.*, **47**, 45 (1978).
- 35) J. Kyte and R. F. Doolittle, *J. Mol. Biol.*, **157**, 105 (1982).
- 36) D. M. Engelman, T. A. Steitz, and A. Goldman, *Ann. Rev. Biophys. Biophys. Chem.*, **15**, 321 (1986).